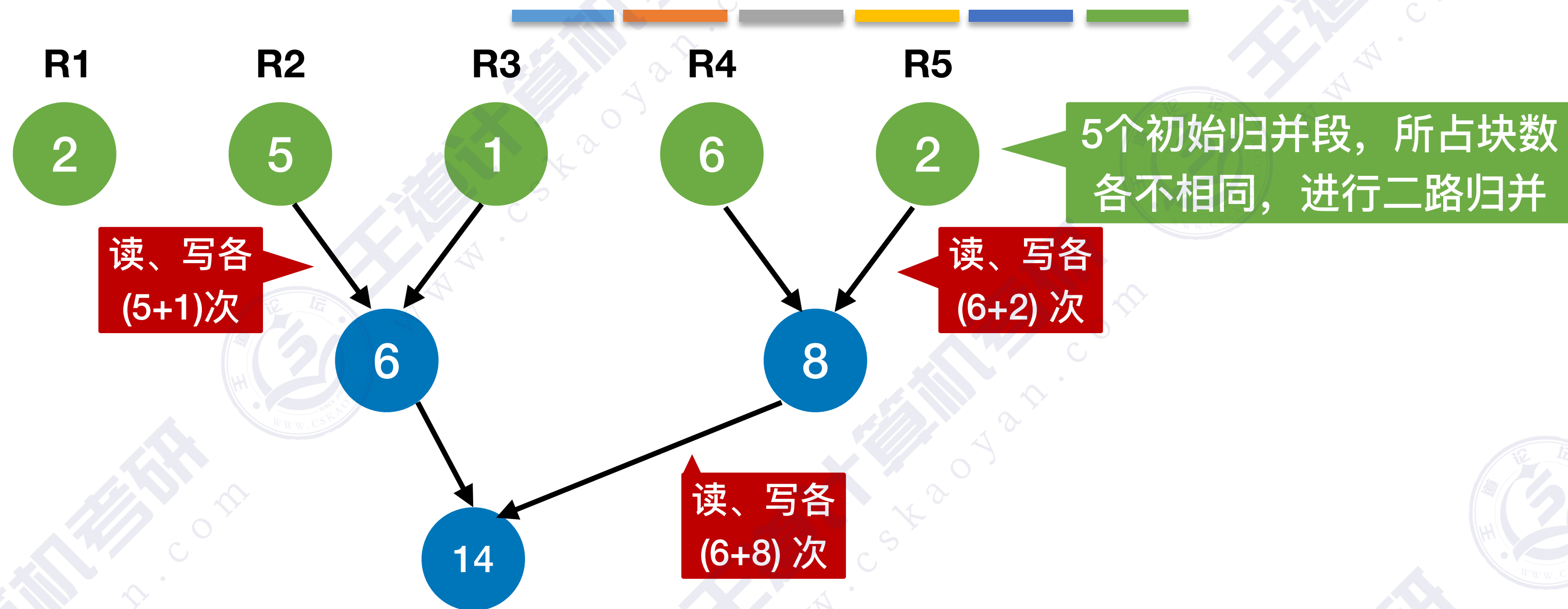


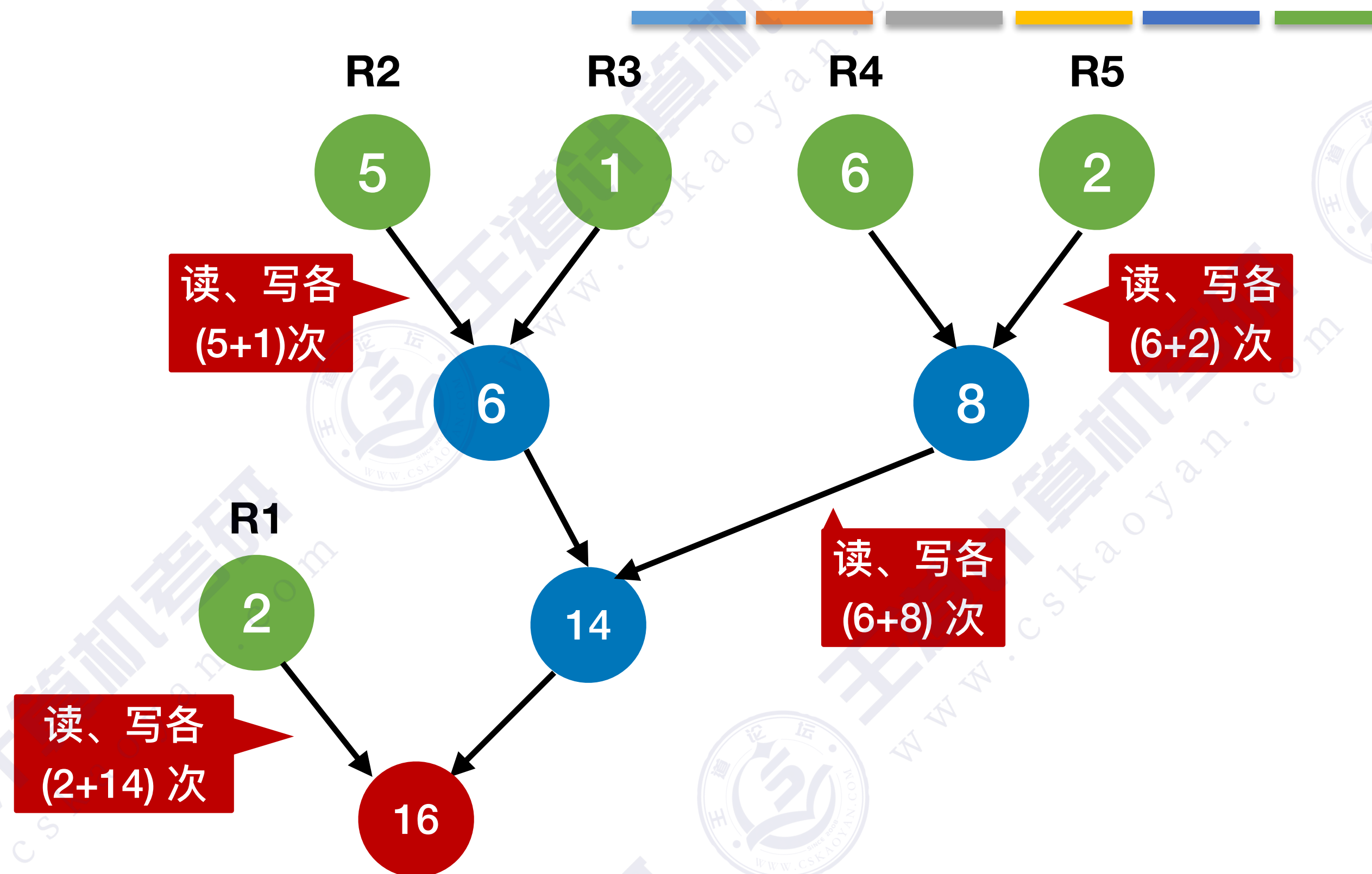
本节内容

最佳归并树

归并树的神秘性质



归并树的神秘性质



要让磁盘I/O次数最少，
就要使归并树WPL最小
——哈夫曼树！

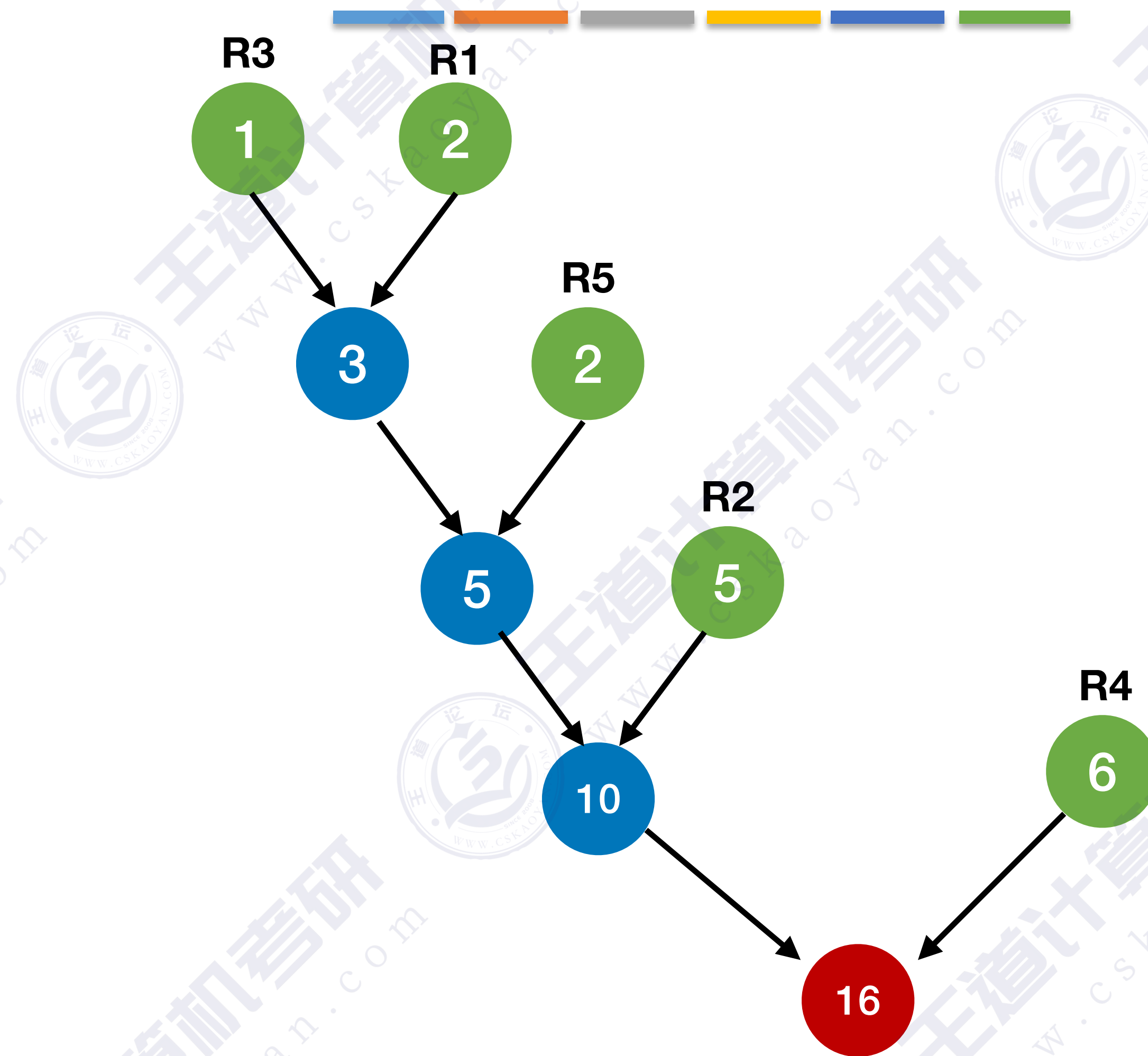


灵光一闪

每个初始归并段看作一个叶子结点，归并段的长度作为结点权值，则上面👉这棵归并树的带权路径长度 $WPL = 2*1 + (5+1+6+2) * 3 = 44 = \text{读磁盘的次数} = \text{写磁盘的次数}$

重要结论：归并过程中的磁盘I/O次数 = 归并树的WPL * 2

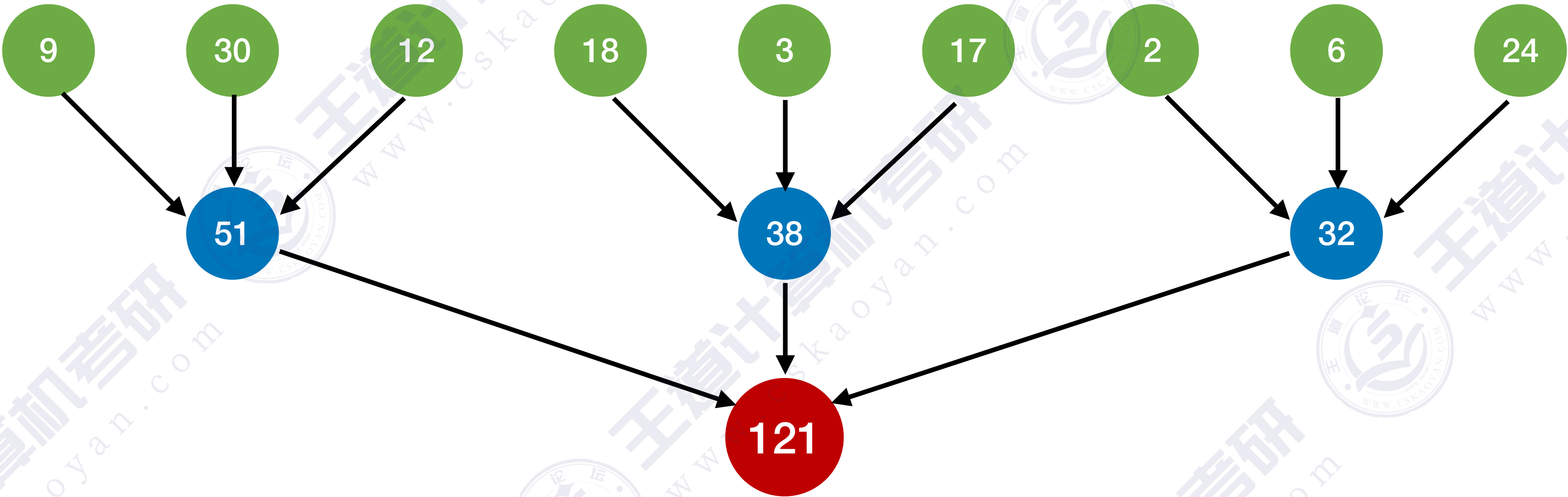
构造2路归并的最佳归并树



最佳归并树 $WPL_{\min} = (1+2)*4 + 2*3 + 5*2 + 6*1 = 34$

读磁盘次数=写磁盘次数=34次；总的磁盘I/O次数 = 68

多路归并的情况



$$WPL = (9+30+12+18+3+17+2+6+24) * 2 = 242$$

归并过程中 磁盘I/O总次数=484次

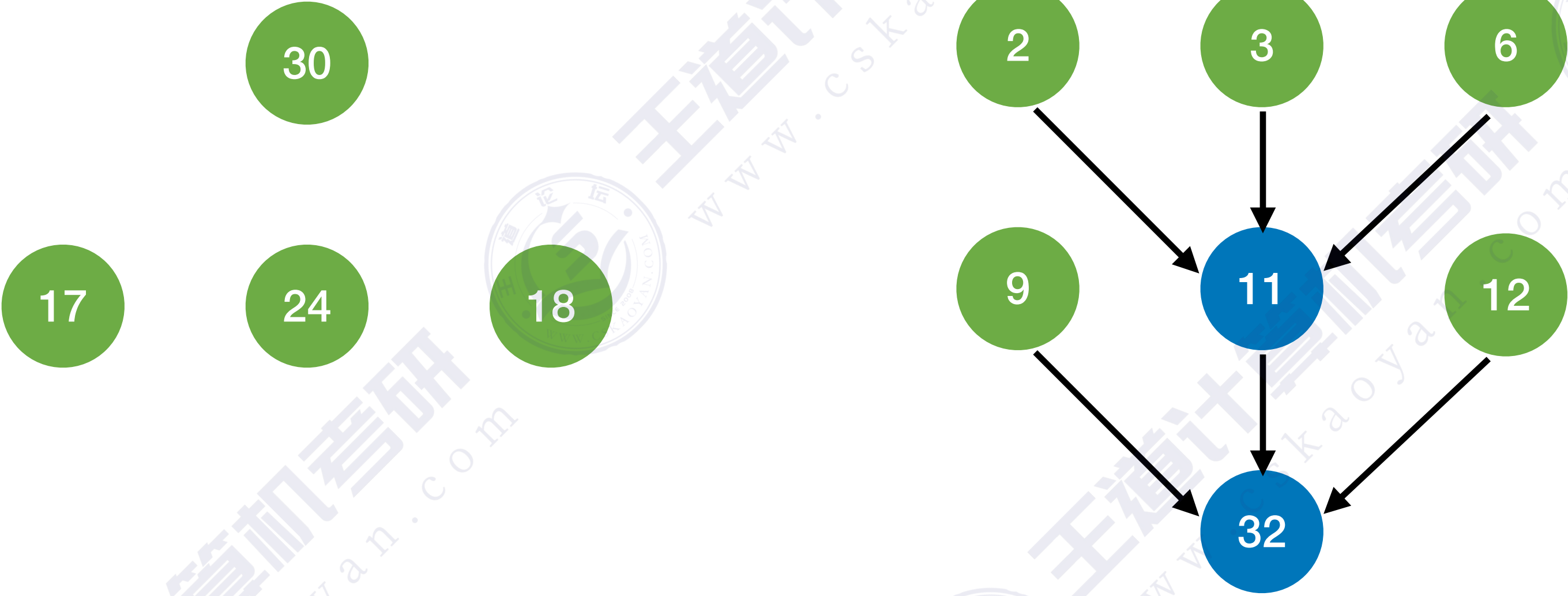
多路归并的最佳归并树



多路归并的最佳归并树

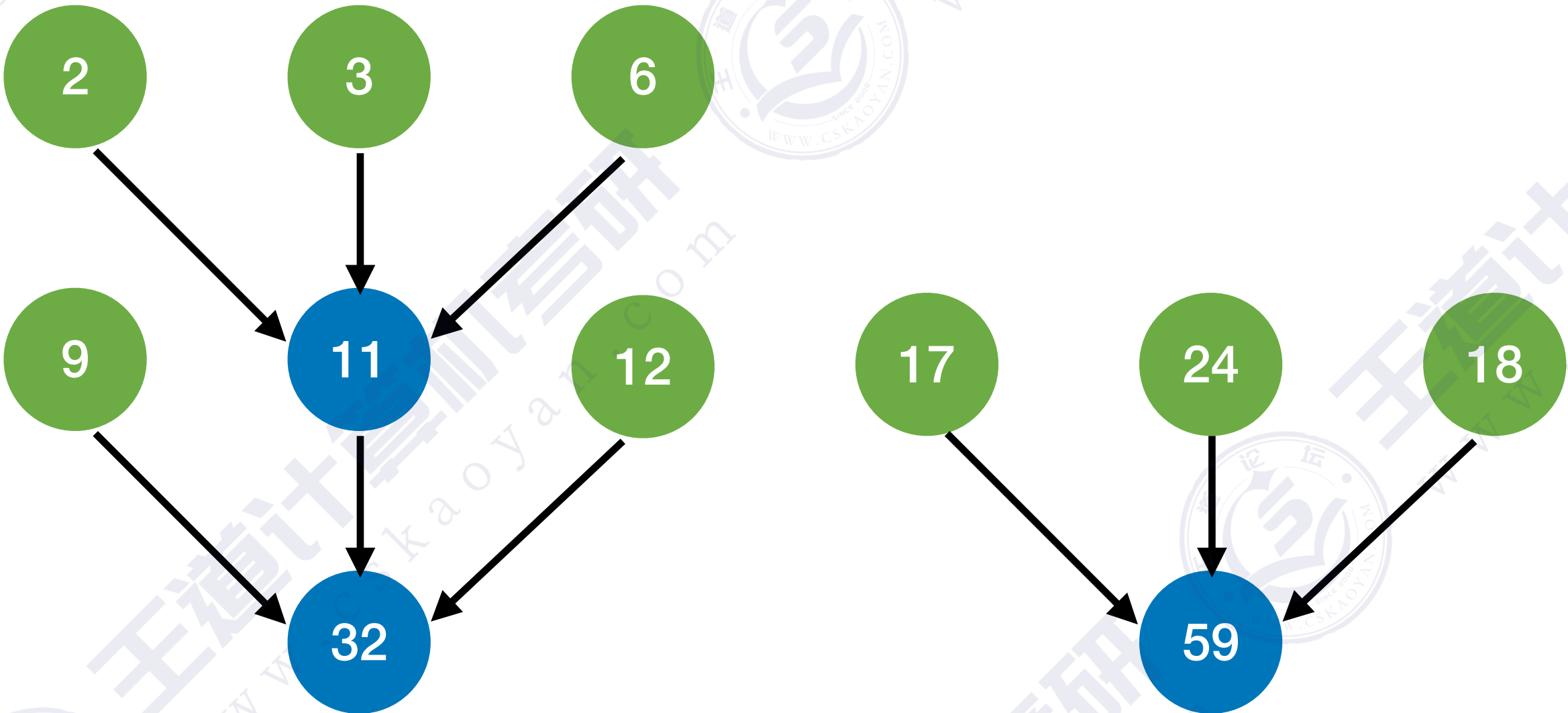


多路归并的最佳归并树



多路归并的最佳归并树

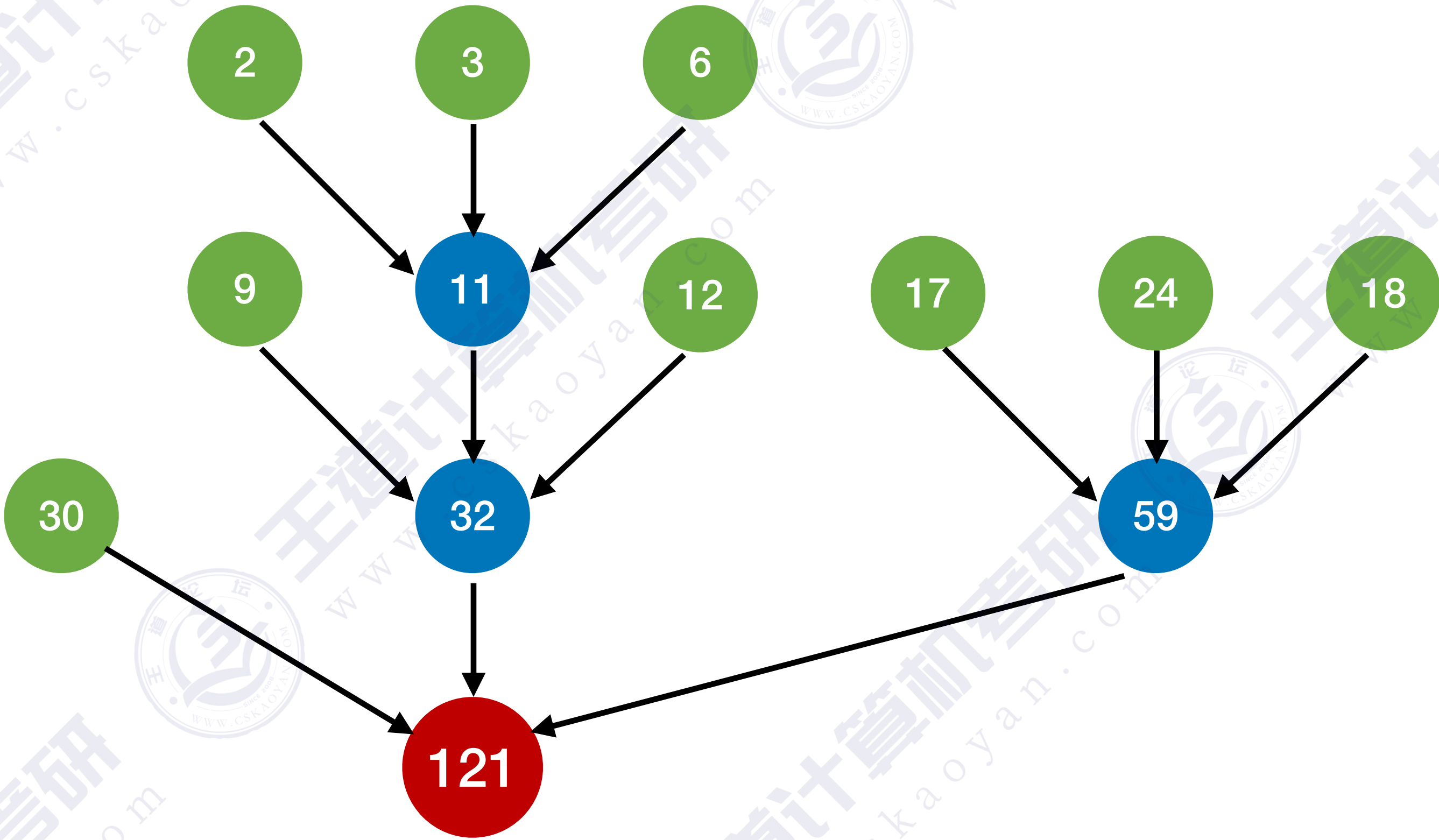
30



多路归并的最佳归并树



右边这棵就是3路归并的最佳归并树



$$WPL_{min} = (2+3+6)*3 + (9+12+17+24+18)*2 + 30*1 = 223$$

归并过程中 磁盘I/O总次数=446次

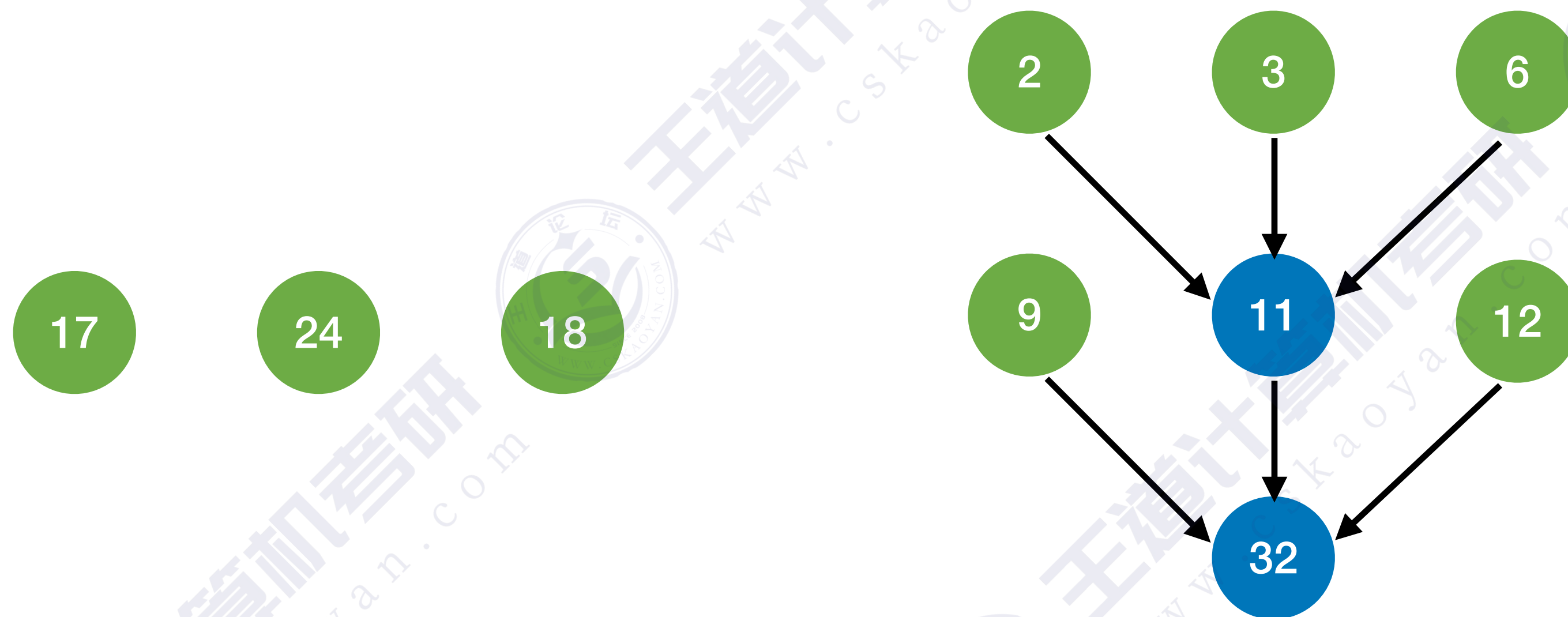
如果减少一个归并段



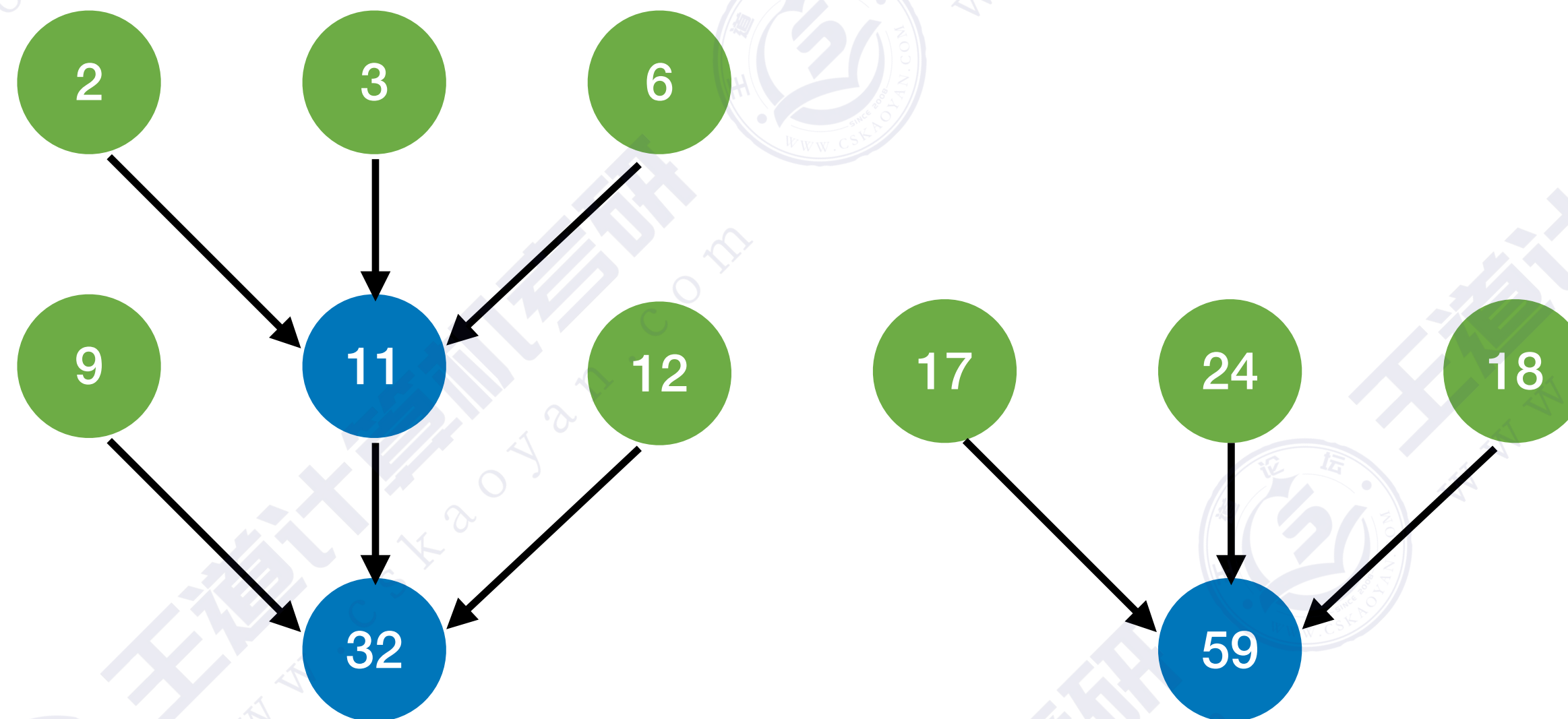
如果减少一个归并段



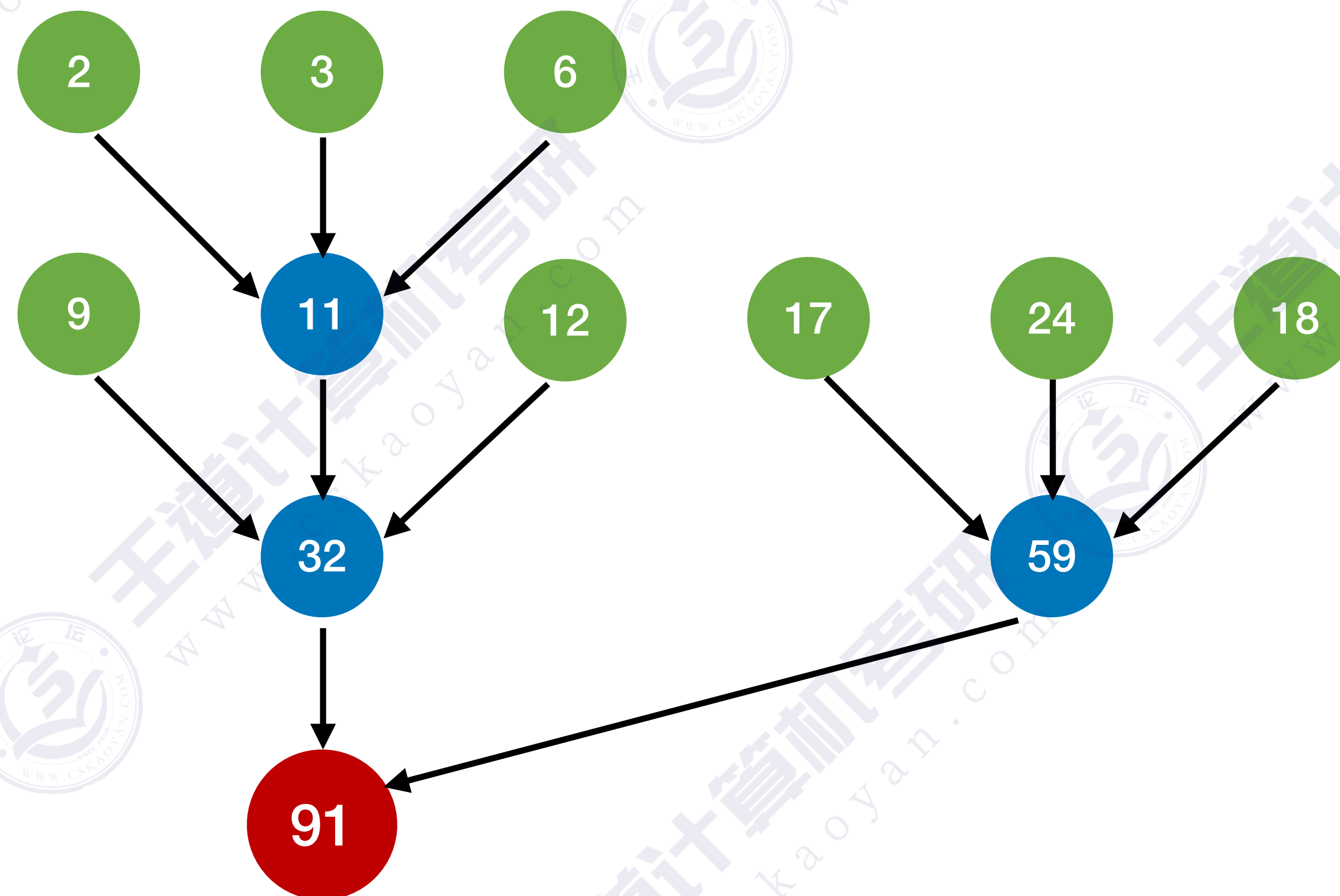
如果减少一个归并段



如果减少一个归并段



如果减少一个归并段



注意：右边这个不是最佳归并树



98k 警告

$$WPL = (2+3+6)*3 + (9+12+17+24+18)*2 = 193$$

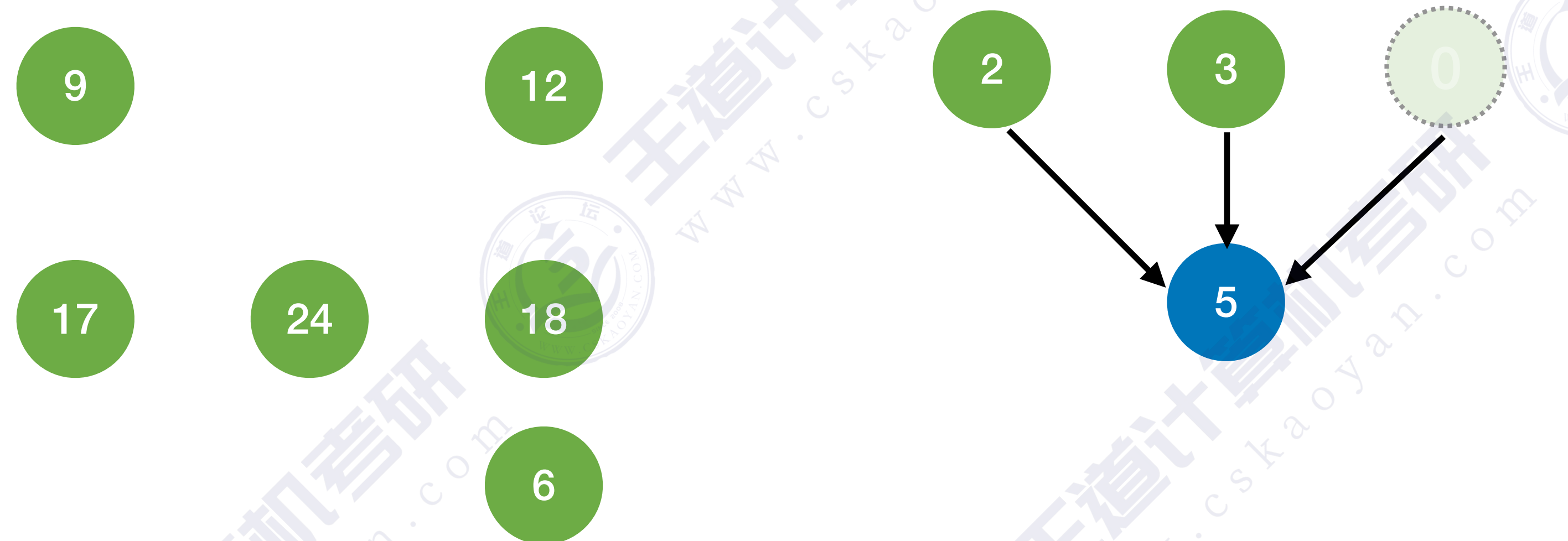
归并过程中 磁盘I/O总次数=386次

正确的做法



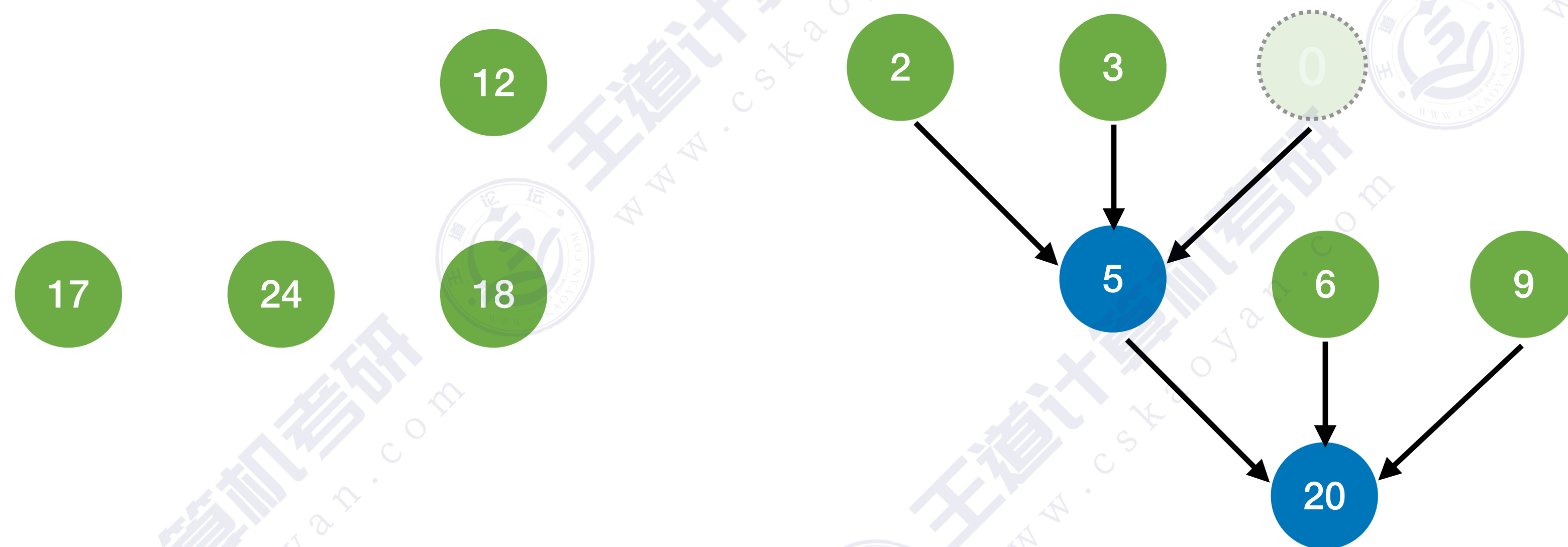
注意：对于k叉归并，若初始归并段的数量无法构成严格的 k 叉归并树，则需要补充几个长度为 0 的“虚段”，再进行 k 叉哈夫曼树的构造。

正确的做法



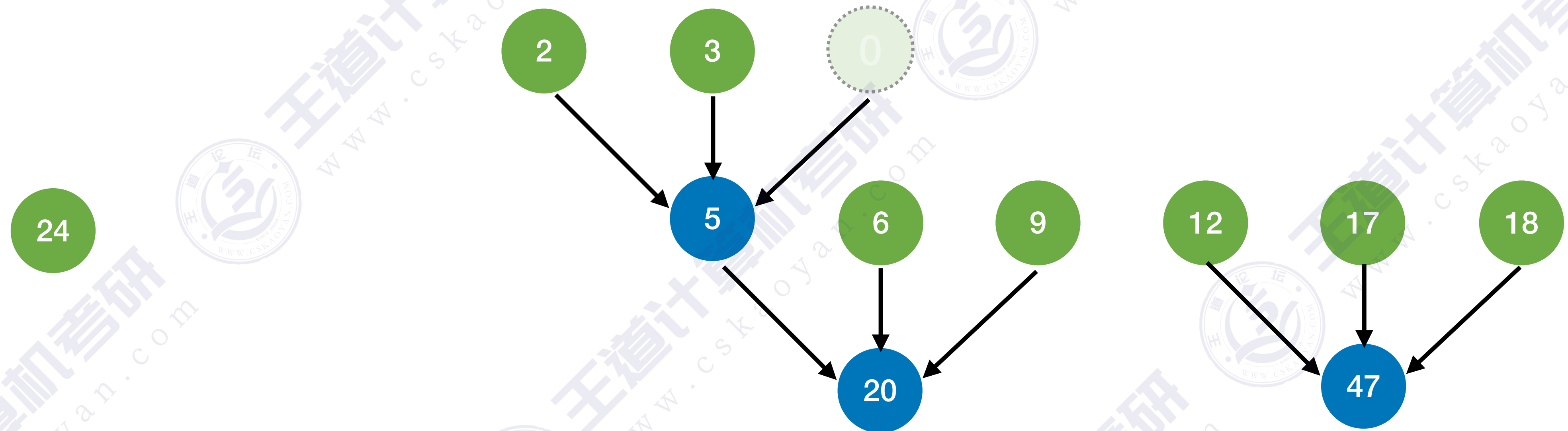
注意：对于k叉归并，若初始归并段的数量无法构成严格的 k 叉归并树，则需要补充几个长度为 0 的“虚段”，再进行 k 叉哈夫曼树的构造。

正确的做法



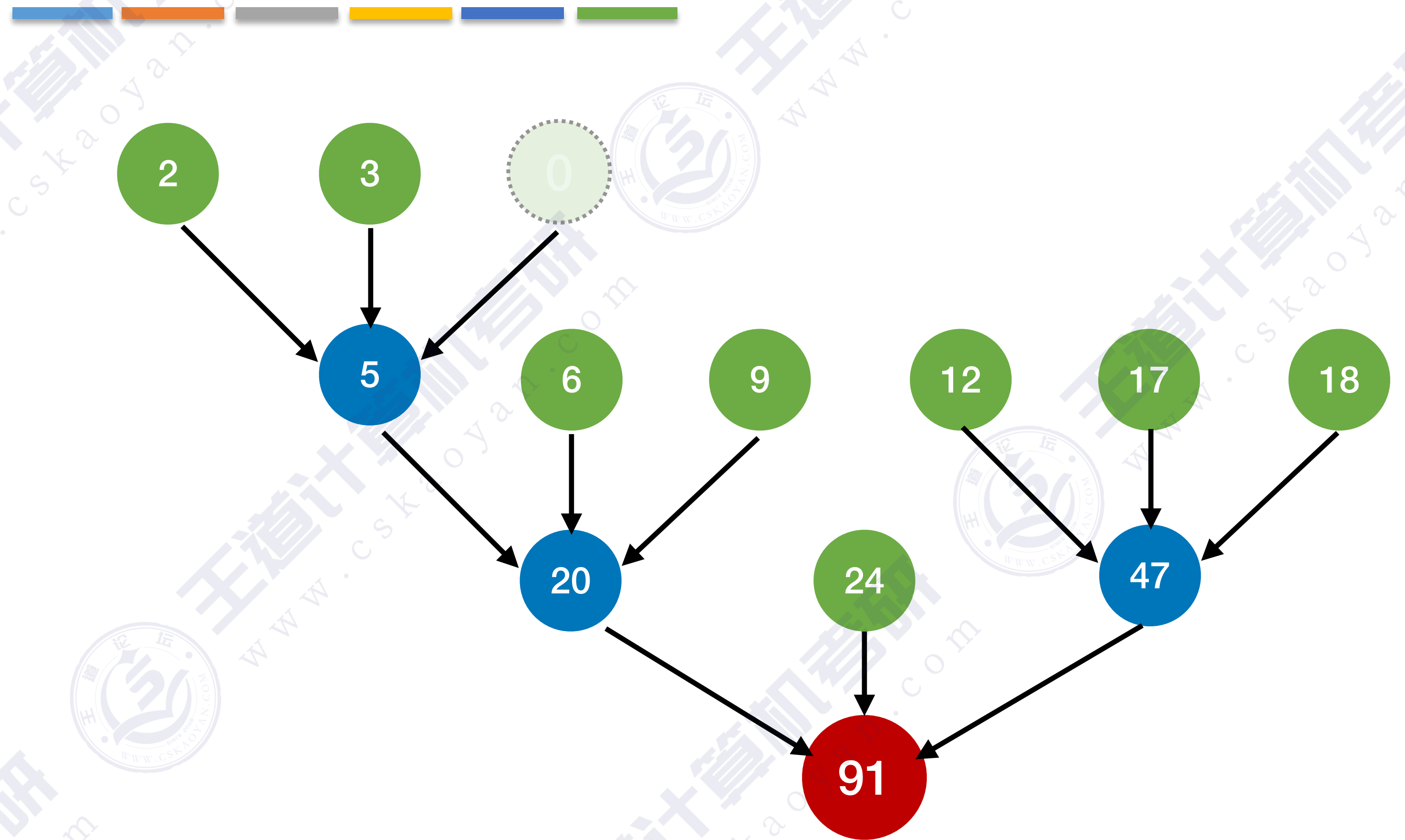
注意：对于k叉归并，若初始归并段的数量无法构成严格的 k 叉归并树，则需要补充几个长度为 0 的“虚段”，再进行 k 叉哈夫曼树的构造。

正确的做法



注意：对于k叉归并，若初始归并段的数量无法构成严格的 k 叉归并树，则需要补充几个长度为 0 的“虚段”，再进行 k 叉哈夫曼树的构造。

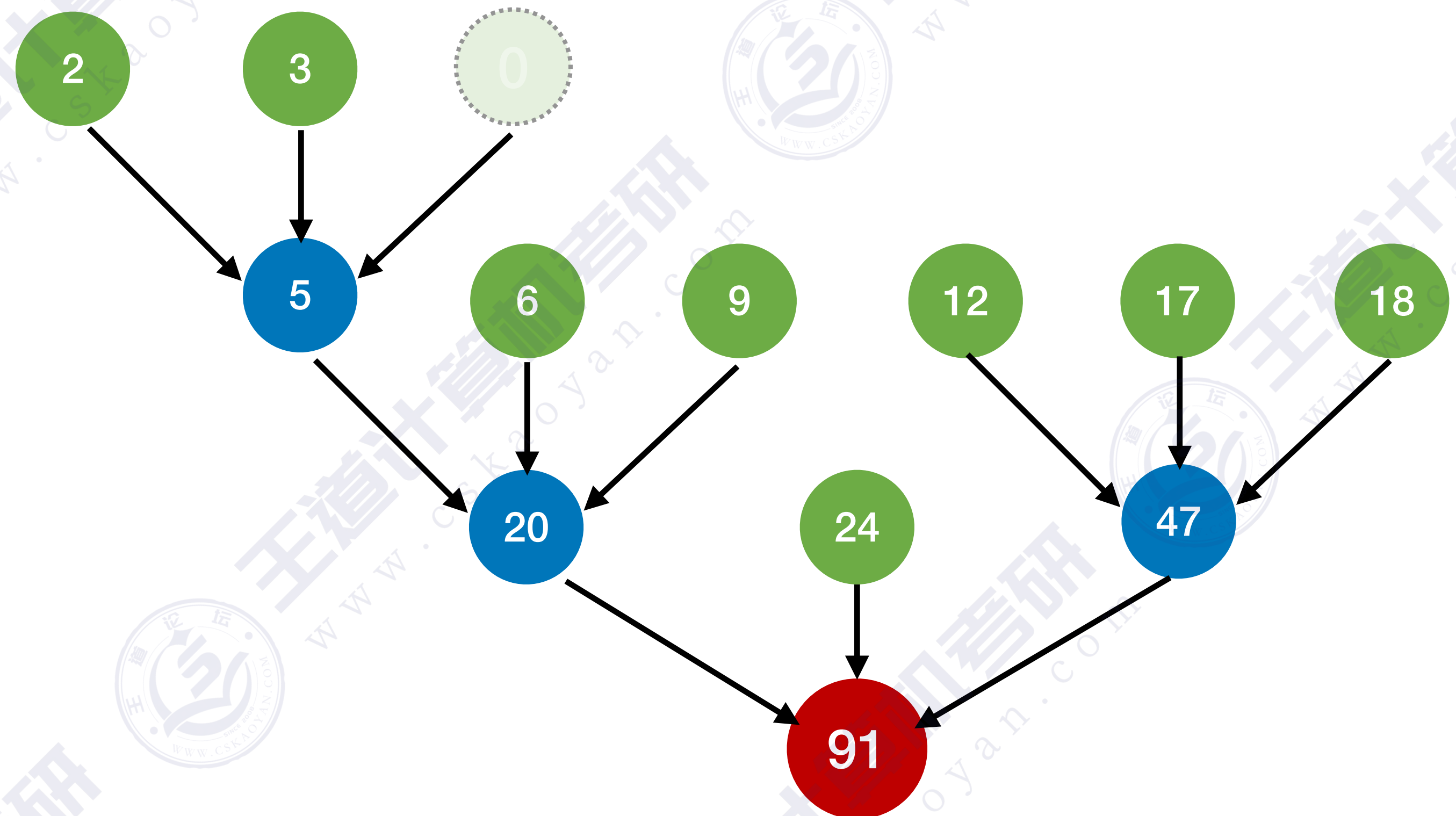
正确的做法



注意：对于k叉归并，若初始归并段的数量无法构成严格的 k 叉归并树，则需要补充几个长度为 0 的“虚段”，再进行 k 叉哈夫曼树的构造。

正确的做法

右边这棵就是3路归并的最佳归并树



$$WPL_{\min} = (2+3+0)*3 + (6+9+12+17+18)*2 + 24*1 = 163$$

归并过程中 磁盘I/O总次数=326次

添加虚段的数量



注意：对于k叉归并，若初始归并段的数量无法构成严格的 k 叉归并树，则需要补充几个长度为 0 的“虚段”，再进行 k 叉哈夫曼树的构造。

老实交代
到底补几个？



k叉的最佳归并树一定是一棵严格的 k 叉树，即树中只包含度为k、度为0 的结点。设度为k的结点有 n_k 个，度为0的结点有 n_0 个，归并树总结点数= n 则：

初始归并段数量+虚段数量= n_0

$$\begin{aligned} n &= n_0 + n_k \\ k n_k &= n - 1 \end{aligned} \quad \longrightarrow \quad n_0 = (k-1)n_k + 1 \quad \longrightarrow \quad n_k = \frac{(n_0 - 1)}{(k-1)}$$

如果是“严格k叉树”，一定能除得尽

- ①若 $(\text{初始归并段数量} - 1) \% (k-1) = 0$ ，说明刚好可以构成严格k叉树，此时不需要添加虚段
- ②若 $(\text{初始归并段数量} - 1) \% (k-1) = u \neq 0$ ，则需要补充 $(k-1) - u$ 个虚段

知识回顾与重要考点

最佳归并树

理论基础

每个初始归并段对应一个叶子结点，把归并段的块数作为叶子的权值

归并树的 WPL = 树中所有叶结点的带权路径长度之和

归并过程中的磁盘 I/O 次数 = 归并树的 WPL * 2

注意：k 叉归并的最佳归并树一定是严格 k 叉树，即树中只有度为 k、度为 0 的结点

如何构造

补充虚段

① 若 $(\text{初始归并段数量} - 1) \% (k - 1) = 0$ ，说明刚好可以构成严格 k 叉树，此时不需要添加虚段

② 若 $(\text{初始归并段数量} - 1) \% (k - 1) = u \neq 0$ ，则需要补充 $(k - 1) - u$ 个虚段

构造 k 叉哈夫曼树

每次选择 k 个根节点权值最小的树合并，并将 k 个根节点的权值之和作为新的根节点的权值

数据结构——剧终

